Standard Sister Clade Comparison Fails when Testing Derived Character States

**JOS KÄFER** AND SYLVAIN MOUSSET

Université de Lyon; Université Lyon 1; CNRS; UMR 5558; Laboratoire de Biométrie et Biologie Evolutive; F-69622, Villeurbanne, France

*Correspondence to be sent to: Laboratoire de Biométrie et Biologie Evolutive; CNRS; UMR 5558; Université Lyon 1; Bât. Grégor Mendel, 43 Bd du 11 novembre 1918, 69622 Villeurbanne cedex; E-mail: jos.kaefer@gmail.com.

Received 13 July 2013; reviews returned 12 November 2013; accepted 19 March 2014

Associate Editor: Luke Harmon

Abstract—Comparing species richness in sister clades that differ in a character state is one of the ways to study factors influencing diversification. While most of its applications have focussed on traits that increase diversification, some have been used to study the association of a trait with lower species richness, e.g., the occurrence of dioecy in flowering plants. We show here, using simulations and an analytical model, that the null expectation of equal species richness that is generally used in sister clade comparisons is wrong in the case of a derived trait occurring independently from speciation: one should expect fewer species in the clade with the derived character state when there is no difference in diversification rates. This is due to the waiting time for the derived state to appear, which causes it to occur more often on longer branches. This has the important implication that the probability for a clade to possess the derived state depends on the tree geometry, and thus on species richness: species-poorer clades are more likely to possess the derived state. We develop a statistical test for sister clade comparisons to study the effect of a derived character state. Applying it to a data set of dioecious clades, we find that we cannot confirm earlier work that concluded that dioecy decreases diversification; on the contrary, it seems to be associated to higher species richness than expected.

11 novembre 1918, 69622 Villeurbanne cedex; E-mail: jos.kaefer@gmail.com.

17 Juillet 2013; reçu en revue le 12 Novembre 2013; accepté le 19 Mars 2014

Associate Editor: Luke Harmon

Abstract—Comparing species richness in sister clades that differ in a character state is one of the ways to study factors influencing diversification. While most of its applications have focussed on traits that increase diversification, some have been used to study the association of a trait with lower species richness, e.g., the occurrence of dioecy in flowering plants. We show here, using simulations and an analytical model, that the null expectation of equal species richness that is generally used in sister clade comparisons is wrong in the case of a derived trait occurring independently from speciation: one should expect fewer species in the clade with the derived character state when there is no difference in diversification rates. This is due to the waiting time for the derived state to appear, which causes it to occur more often on longer branches. This has the important implication that the probability for a clade to possess the derived state depends on the tree geometry, and thus on species richness: species-poorer clades are more likely to possess the derived state. We develop a statistical test for sister clade comparisons to study the effect of a derived character state. Applying it to a data set of dioecious clades, we find that we cannot confirm earlier work that concluded that dioecy decreases diversification; on the contrary, it seems to be associated to higher species richness than expected.

© The Author(s) 2014. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.


What explains the extraordinary diversity of life in general, and certain clades in particular, is one of the major questions in evolutionary biology. Some characters have been identified as “key innovations” (e.g., Hunter 1998), when the appearance of this character leads to an increase in diversification. The opposite also exists: when a trait that has evolved diminishes the species richness of the clade with which it is associated, it is said to be a macroevolutionary “dead end” (Wiegmann et al. 1993; Heilbuth 2000; Escober et al. 2010; Hardy and Cook 2010).

The tool often used to perform such studies is a comparison of sister clades in a phylogeny (e.g., Slowinski and Guyer 1993; Barracough et al. 1998; Vamosi and Vamosi 2005; Ricklefs 2007), which is supposed to be more reliable than comparing groups on the basis of taxonomy (Isaac et al. 2003). In such tests, one counts the number of species in two clades that have the same common ancestor but differ in the state of the character under study. Applied to a sufficiently large number of sister pairs, one can then test for differences in diversification against the null hypothesis that both clades do not differ on average in their size. Sister clade comparisons are considered powerful because they circumvent problems due to differences in clade age, or taxonomic biases. They allow one to use large numbers of species and groups, which is not always possible with more recent methods (e.g., BiSSE, FitzJohn et al. 2009) that require detailed and time-calibrated phylogenies.

Several improvements of sister standard clade comparison tests have been developed, notably based on maximum likelihood methods (e.g., McConway and Sims 2004; Paradis 2012). These tests have been used repeatedly during the past 20 years to study diversification in a variety of groups, such as insects (Wiegmann et al. 1993; Davis et al. 2010; Hardy and Cook 2010), birds (Mooers and Moller 1996), and plants (Davies et al. 2004; Vamosi and Vamosi 2010, 2011).

Dioecy (i.e., having individuals of different sexes) occurs in only 5–10 % of the flowering plants species (Renner and Ricklefs 1995; Vamosi et al. 2005), and has evolved independently on many different occasions (Lebel-Hardenack and Grant 1997). Heilbuth (2000) used sister clade comparisons to study the low occurrence of dioecy in flowering plants could be related to higher extinction and/or lower speciation rates, leading to lower diversification. She found that in a statistically significant majority of the comparisons, dioecious clades are smaller than their non-dioecious sister clades, and concluded that dioecious species diversify less. Kay et al. (2006), using similar methods but a different, smaller data set, came to similar conclusions. These results inspired later work in which hypotheses that explain why dioecy is not advantageous were developed (Heilbuth et al. 2001; Vamosi and Otto 2002), reviewed (e.g., Barrett 2002), and tested (e.g., Vamosi and Vamosi 2004; Queenborough et al. 2009), including a recent paper by one us (Käfer et al. 2013).

We here argue and show that the null expectation of having equal clade sizes in sister clade comparisons does not hold if one of the character states is a derived state, such as dioecy, and if transitions occur independently from speciation (anagenetic transitions, Stuessy et al. 2006). In fact, the number of species in the clade with the derived state will on average be smaller than in the clade with the ancestral state. This result has implications for sister clade comparisons, and should be taken into account in any comparison using species richness estimates. We develop a method that can be applied to a sister clade dataset, and show that the conclusion that dioecy decreases diversification was premature, as the opposite might be true.
First, we consider a tree in which speciation has been constant throughout evolution, and in which there is no extinction. Thus, each branch, at any point in time, has had the same probability to split into two. Transitions from the ancestral to the derived state occur with a constant rate throughout the history of the tree. In such a random process, the longer the branch, the larger the probability that the derived state has appeared on it. If one selects sister clades on such a tree that differs in the trait state, one will end up with the derived clades having longer stem branches than their sisters on average. Thus, the crown groups of the derived clades will be younger than the crown groups of the sisters on an average, and with equal diversification rates, they will contain fewer species, on average.

In other words, while it is true that sister clades have one ancestor in common and are thus equally old, the criterion that we chose to select them (contrasting in character state) is not of the same age as the ancestor—it is of more recent origin. The derived clade only had time to diversify from this transition on, while the sister clade could diversify from the ancestral node on. Therefore, the clades are not expected to have the same species richness.

**Mathematical Model**

If the difference in stem branch length of the two crown groups, assuming equal diversification rates, determines their difference in species richness, then it should be possible to estimate the stem branch length difference as a function of the species richness differences.

We use the Yule (pure birth) model of speciation (Yule 1924) with speciation rate \( \lambda \), and assume that a unique and irreversible (Dollo 1893) transition from an ancestral to a derived state occurs on one stem branch (see Figure 1) of the species tree; transition events are assumed to be anagenetic. The probability for the transition to occur on a branch is proportional to its length. The total number of species in the tree is denoted \( M \) and will be treated as a parameter. The initial speciation event partitions the \( M \) species into two clades. We arbitrarily choose to orientate the tree such that the second speciation event occurs on the left stem branch. The number of species in the clade descending from the right stem branch is denoted \( K \) and verifies \( 1 \leq K \leq M-2 \) because of the arbitrary orientation of the tree.

Equation [1] in the appendix gives the conditional distribution of \( K \) given \( M \). The probability for the left (shorter) stem branch to have \( p \) descendants is \( P(K=p|M=m) \); the probability for the right (longer) stem branch to have \( p \) descendants is \( P(K=p|M=m) \).

The probability density of the length of a stem branch in a Yule tree with \( M \) extant species has been studied (Stadler and Steel 2012), however the conditional probability density given the clade sizes in an oriented tree is unknown.

First, we consider a given bipartition of \( p \) and \( m-p \) species. The probability of a given bipartition can be deduced from Equation [1] in the Appendix:

\[
P(K \in \{p, m-p\}|M=m) = \begin{cases} \frac{2}{m^2}, & \text{if } p \neq \frac{m}{2} \\ \frac{m}{2}, & \text{if } p = \frac{m}{2} \end{cases}
\]

which is a uniform distribution with the sole exception of \( p = \frac{m}{2} \), which has one half the probability of any other split (Farris 1976; Purvis et al. 2002).

The conditional distribution of the number of species \( D \) in the derived clade is then given by the relative lengths of the stem branches, which we propose to approximate by the relative expectations of the stem branch lengths

\[
P(D=p|M=m, K \in \{p, m-p\}) \approx \begin{cases} \frac{2}{m^2} \mathbb{E}(L_{m-p}), & \text{if } p \neq \frac{m}{2} \\ 1, & \text{if } p = \frac{m}{2} \end{cases}
\]

(1)
where $\mathbb{E}(L_m,p)$ (respectively $\mathbb{E}(L_m,m-p)$) is the expected length of the stem branch with $p$ (respectively $m-p$) descendants (Equation [4] in the Appendix). This approximation relies on the assumption that the probability of a state change is proportional to the branch length, and thus holds as long as the transition rate is low relative to the speciation rate.

Assuming all partitions are equally likely to harbor a transition event and thus to be sampled in a sister clade analysis, the conditional distribution of the number of species in the derived clade is given by:

$$P(D=p|M=m) = P(D=p|M=m,K\in\{p,m-p\}) \times P(K\in\{p,m-p\}|M=m).$$

(2)

Computer simulations were used to test the reliability of the approximation in Equation [1]: Yule trees with a fixed number of species $M$ were generated; for each tree one of the stem branches was randomly chosen (with a probability proportional to their relative lengths) to host a transition event and the size of the derived clade $D$ was computed. Figure 2a shows the resulting distribution of $D$ for $M=50$.

However, assuming that all partitions are equally likely is unrealistic, since the probability for a clade pair to harbor a transition event depends on the lengths of the stem branches, leading to a selection bias for imbalanced partitions with relatively longer stem branches. The conditional sampling probability of a bipartition given the total number of species $M$ thus depends on the lengths of its stem branches relative to the lengths of the stem branches of other possible bipartitions. Whatever the value of $M$, the left (shorter) stem branch ends at the second speciation event and its length is exponentially distributed with mean $1/\lambda$. The expected length of the right (longer) stem branch given $K$ and $M$ ($\mathbb{E}(L_{m,k})$) is given by Equation [3] in the Appendix. We propose to approximate the conditional distribution of the number of species in the derived clade $D$ given (1) the total number of species $M$ and (2) that a transition has occurred on one of the two stem branches, by

$$P(D=d|M=m) \approx \frac{1}{\lambda} \times P(K=m-d|M=m) + \mathbb{E}(L_{m,k}) \times P(K=k|M=m) \sum_{k=1}^{m-1} \left( \frac{1}{\lambda} + \mathbb{E}(L_{m,k}) \right) \times P(K=k|M=m).$$

(3)

where the speciation rate $\lambda$ cancels out. Computer simulations showed that this approximation is reliable: the theoretical and simulated distributions of $D$ in pairs of clade with $M=50$ are plotted in Figure 2b. The mathematical model is based on a pure birth (Yule) model, but can be directly compared with these simulations that were performed using BiSSE (described in more detail in the section “Simulations”) with a nonzero extinction rate.

**Figure 2.** Theoretical and simulated distributions of the number of species in the derived clade in an anagenetic transition model (solid black line) compared to the uniform distribution that underlies previous sister clade comparison tests (gray lines). (a) Without correction for selection bias: theoretical prediction using Equation [2] (solid black line). Simulations of 100,000 Yule trees with 50 species (circles) and 95% prediction interval for 100,000 simulations (gray-shaded area) of the uniform distribution. (b) Corrected for selection bias, given a phylogenetic process with constant diversification rate. Theoretical prediction using Equation [1] (solid black line). BiSSE simulations of trees with 50 species, retaining only trees in which a transition occurred on one of the stem branches. Parameters are $\lambda_d=\lambda_s=0.1$, $\mu_d=0.04$, $T_{d-c}=0$, and $T_{s-c}=0.004$ (circles) or $T_{s-c}=0.0004$ (triangles). For both combinations of parameters, simulations were performed until 3,000 trees were obtained. 95% prediction interval for these simulations (dashed black lines) and for the uniform distribution (gray-shaded area) for 3,000 simulations.
A Statistical Test

While the selection bias can be modeled in the case of constant diversification rates (Equation [3]), the application to real data will be difficult, as assuming constant diversification rates is a very strong assumption. Indeed, diversification rates are shown to vary (cf. Magallon and Sanderson 2001 for angiosperms), and this will increase the probability of finding imbalanced bipartitions (Farris 1976, and see also Scotland and Sanderson 2004 for an illustration of the problem of simulating a clade size distribution that corresponds to observed distributions). Furthermore, the application of Equation [3] in a classical hypothesis testing framework, like the method used by Slowinski and Guyer (1993), would be very computer-intensive, as all of the bipartitions need to be taken into account.

To circumvent these two problems, we propose to use a resampling method that conditions on the c sampled bipartitions and reassigns the states of the sister clades based on the ratio of the expected stem branch lengths given the observed species richness of sister clades. Thus, we “forget” which one of the sister clades is in the derived state, and randomly assign the states according to the probability given by Equation [1]. The resulting resampled clade pair has the same species richness as the observed one, but the states might have been swapped.

We create a large number of data sets of c resampled clade pairs, and use the mean proportion of species in the derived clade among the sampled pairs, \( \frac{d}{m} = \frac{1}{m} \sum_{i=1}^{m} \frac{d}{m} \), as a test statistic. We use the following algorithm:

1. For the i-th clade pair, the probability for the transition to have occurred on the stem branch of the observed derived clade is
   \[ p = \frac{\mu_{d} \mu_{s}}{\mu_{s} \mu_{d} + \mu_{s} \mu_{a} + \mu_{d} \mu_{a}} \] (Equation [1]).

2. The resampled numbers of species in the derived clades \( d' \) are reassigned for the c clade pairs as follows:
   \[ \frac{d'}{m} = \begin{cases} d, & \text{with probability } p; \\ m - d, & \text{with probability } 1 - p. \end{cases} \]

3. \( \frac{d}{m} \) is computed from the resampled data set.

4. The process is repeated from step 2 a large number of times to obtain a simulated null distribution of the \( \frac{d}{m} \) statistic.

The observed \( \frac{d}{m} \) value is then compared to the simulated null distribution in a classical hypothesis testing framework.

R (R Development Core Team 2011) source code to perform this test is available as Supplementary Material at the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.jdfvg. It computes the \( p \)-value associated to one- or two-sided alternative hypotheses from a list of values of \( m \) and \( d \).

Simulations

We tested the model and the statistical test in simulated phylogenetic trees, in which simultaneous speciations and extinctions can occur anywhere with a fixed probability, and with a nonzero extinction rate. This situation is closer to reality, where one selects sister clades from a phylogeny, than the mathematical model described above that assumes the sister clades to be given.

For this purpose, we used the tree simulator included in the BiSSE (Binary State Speciation, and Extinction) software (Maddison et al. 2007; FitzJohn et al. 2009); starting with the ancestral state, speciations (branch splitting), and extinctions (branch termination), occur with rate \( \lambda_d \) and \( \mu_d \), respectively, and transitions to the derived state with rate \( T_{d\rightarrow d} \). Equally, for the derived state, \( \lambda_d \) and \( \mu_d \) are the speciation and extinction rates respectively, and \( T_{d\rightarrow a} \) is the reversion rate, which corresponds to the loss of the derived state. The diversification rate for a trait is simply \( \lambda - \mu \).

BiSSE is implemented in the R (R Development Core Team 2011) package “diversitree,” and the simulated trees were analyzed with functions from the “ape” package (Paradis et al. 2004). We simulated trees with 30,000 species each, with different combinations of the parameters as shown in Table 1. We counted, for each simulation, the number of sister clade pairs \( c \), and for each pair \( i \) the number of species \( m_i \) of species in both clades, the number \( d_i \) of species in the clade in the derived trait, and the ratio \( \frac{d_i}{m_i} \).

First, we considered equal diversification rates by fixing the extinction rates to be equal, and varied the values of \( T_{d\rightarrow d} \) to influence the number of species with the derived trait (first five rows of Table 1). Indeed, we observe that the clade in the ancestral state is on average larger than the one with in derived state, and the average ratio \( \frac{d}{m} \approx 0.42 \).

Using these trees that were simulated under the null hypothesis, we studied the actual type I error rate of the resampling test we propose here. We computed the actual rejection rate of the one-tailed test aimed to detect larger species numbers in the derived clade and compared this rejection rate to the critical value of \( a = 5\% \) in a one-sided exact binomial test. None of these five series of simulations showed evidence for an increased actual rejection rate, enabling us to think that the resampling test is rather conservative. Taking these five simulation series together, the null hypothesis was rejected in 76 out of 1261 trees, which is slightly higher than 5%, but not significantly according to a one-sided exact binomial test (\( p \)-value=0.057).

Reversions from the derived to the ancestral state can occur, and in these cases, the derived state should be considered the ancestral state of the clades pair. However, in practice, clade pairs will occur for which

\[ M = \sum_{i=1}^{m} d_i \]

This means that the expected number of species in the derived clade increases with the number of species in the ancestral clade.
it will be difficult to determine which state was ancestral. We performed a series of simulations allowing increasing reversion rates (rows six to ten Table 1). This does not seem to lead consistently to a large increase in the type I error rate. First, because reversions can only take place after a transition, they will generally be rare, even if the transition and reversion rates are equal. Second, we chose to count each species in the derived state only once: if several clade pairs can be assembled into a larger group, this is an indication of the fact that reversions have occurred, and that only the largest group (including all subsequent smaller clade pairs) should be counted as a clade pair in the sister clade comparison.

Then, we studied the power of the resampling test under alternative hypotheses where the derived state has an influence on the diversification rate $\lambda - \mu$. To study the case of a higher diversification rate associated to the ancestral state, we either increased the extinction rate of derived branches, or increased the speciation rate of the ancestral ones (Table 1, row 11 and 12). The relative species richness of the derived clades decreases moderately, but sufficiently to be detected with the sister clade comparison test to reject the null hypothesis of lower species richness in the derived clade depending on the difference in diversification rates (about 25% for a difference of 0.03, and 70% for a difference of 0.1). When the derived state was associated to a higher diversification rate (by increasing extinction of the ancestral state, or increasing the speciation of the derived state, Table 1, rows 13 and 14), the value of $|d|$ increased, but remained below 0.5. The relatively high value of $|d|$ compared to $|m-d|$ indicates that large clades in the derived state do occur, yet these are relatively rare. The older the clade, the smaller the difference in branch lengths of the sisters will be relatively to the total clade’s age. Thus, we would expect the few old derived clades to be larger, while the majority of the younger derived clades will be smaller than their sisters. The resampling test (one-tailed) has reasonable power to detect this difference in diversification.

To illustrate what would happen if one applied the classical sister clade comparison tests to anagenetic transitions, we used several available tests on the same simulation results. A test known as the diversity contrast test (e.g., Vamosi and Vamosi 2005) uses a Wilcoxon test on some ratio of the species richness of the sister clades; we here chose the log-transformation of the species richness (Barracoulough et al. 1996; Vamosi and Vamosi 2005; Paradis 2012). The sign test (e.g., Heilbuth 2000; Vamosi and Vamosi 2005) relies on counting the
number of times the derived clade is larger than its sister. McConway and Sims (2004) propose a version of the test by Slowinski and Guyer (1993) using a maximum likelihood approach. Finally, a recently proposed test by Paradis (2012) also uses a maximum likelihood approach, but was designed to deal with some of the shortcomings of the test by McConway and Sims.

All these tests rely on the assumption that, under equal diversification rates, both sisters should be drawn from the same distribution of sizes. Not surprisingly, as we show in Table 1, the type I error rate for these tests is thus very high. In a majority of the simulated trees with equal diversification rates, the null hypothesis is rejected with the diversity contrast test, the sign test, and the test by Paradis (2012); sometimes the rejection rate is close to 100%. McConway and Sims’ test, on the contrary, has a lack of power. Its low rate of rejection of the null hypothesis in the first ten rows is however mainly due to its sensitivity to small clade sizes, as already noted by McConway and Sims.

Some of these tests, especially those that are not based on maximum likelihood testing, could easily use the findings we present here to change the null hypothesis. However, a test based on resampling statistics is the most straightforward way to deal with the fact that the selection of sister clades is a nonrandom sampling scheme, and guards against violations of implicit assumptions.

APPLICATION TO DIOECY IN ANGIOSPERMS

We applied this test on the dataset of dioecious clades compiled by Heilbuth (2000). She divided her data in two sets, one on the family level and one on the genus level, and reached the same conclusion for both data sets. As sister group comparisons do not need to be restricted to a taxonomic level (it might even be desirable that they are not), we merged them into one, suppressing nearly identical entries due to monogeneric families.

As the phylogeny of some of the clades was uncertain, Heilbuth was not always able to unambiguously assign a sister clade to the dioecious clade. To deal with those cases, we used two data sets, using either the smallest or the largest possible non-dioecious sister clade. In total, we used 89 pairs of sister clades, of which 9 had an ambiguous phylogenetic position. For the data set with the smallest possible sister clades, \( \frac{m}{d} = 0.35 \), while when the smallest possible sisters were chosen, \( \frac{m}{d} = 0.37 \).

We resampled the data by reassigning the derived state to one of the sisters in each pair according to their estimated stem branch lengths (Equation [1]), as described above. Based on the data set with the largest possible sister clades, in 100,000 resampled sets, 2,127 had a \( \frac{m}{d} \)-value greater than the observed value. Using Heilbuth’s original (nonconservative) sister clade sizes leads to similar results (1932 out of 100,000 simulated \( \frac{m}{d} \)-values greater than the observed value). The p-value of 0.0213 for the most conservative data set would allow us to accept the alternative hypothesis that dioecious clades are more species-rich than expected under equal diversification rates, which is the contrary of the conclusion Heilbuth obtained with classical sister clade tests.

The data set and code to readily apply the resampling test to these data are available as Supplementary Material at the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.jbhv3.

DISCUSSION

Sister Clade Comparisons

Sister clade comparisons have a long history (e.g., Slowinski and Guyer 1993; Wiegmann et al. 1993; Mooers and Moller 1996; Heilbuth 2000; Davies et al. 2004; Davis et al. 2010), and their use has been advocated as being a way to avoid problems due to nonmonophyly, dependence, and differences in taxon age (Barraclough et al. 1998; Isaac et al. 2003). We here discovered a bias in sister clade comparisons which, to our knowledge, has previously passed unnoticed: if one compares clades that present a derived state with their sisters that possess the ancestral character state, one should expect, under equal diversification rates, the clade with the derived state to be on average smaller than the one in the ancestral state.

The cause for this bias is the waiting time for the derived character state to appear after the split into two sister clades: this character state could have appeared anywhere on the stem branch, thus the longer this branch, the higher the probability that the clade is in the derived state. As the clade with the longer stem branch has the youngest crown age in a pair of sister clades, this automatically implies that it will have, on average, fewer species. Thus, the null hypothesis of a test based on sister clade comparisons should be that the clades with the derived state are to some extent smaller than their sisters. Most sister clade comparisons have been performed to detect whether a character state increased diversification, and our findings make these tests conservative: the old null hypothesis of equal species richness is, according to our results, already an indication for increased diversification associated with the derived character state.

More generally, the probability for a clade to possess the derived character state depends on the clade geometry, and accordingly its species richness. If transitions occur anywhere in the tree with equal probability, long branches will “attract” these transitions more often than short ones. Even under equal diversification rates, the stochastic nature of the branching process is such that large differences in species richness between clades are easily observed (Farris 1976). This situation could become even worse when diversification rates vary, which they appear to do in the angiosperms (Magallon and Sanderson 2001; Davies
that the transition rates are very small compared to the of the dioecious clades. Fourth, our method assumes could be the cause of larger species richness in some real trees (cf. Ricklefs 2007; Rabosky 2009), the chances occur anagenetically, i.e., that do not coincide with a speciation. However, as extinction rates are often large in this result, for four reasons. First, Heilbuth meant to diversification rates, seems to lead to higher species richness using the same data, that dioecy, rather than decreasing this result also invites us to re-address the question what causes the “ecological correlates of dioecy” (Bawa 1980; Renner and Ricklefs 1995; Vamosi et al. 2003). Vamosi et al. (2003) argued that these traits decreased the extinction risk, thus only the dioecious clades presenting these traits would have survived. However, some of these traits, like a woody growth form, and abiotic pollination, have been associated to decreased diversification (Kay and Sargent 2009; Vamosi and Vamosi 2011), and seem not to be ancestral in many angiosperm clades, just like dioecy. Thus, these traits might have occurred on the same long branches as the transition towards dioecy, causing a correlation with dioecy that would be due to the geometry of the phylogenetic tree.

**Perspectives**

A prediction that can be made from these results, and which could be used to test their generality, is that clades that harbor the derived state have generally younger crown ages than their sister clades in the ancestral state. Once crown ages are known, more precise estimations of diversification rates could be made. However, for a large number of clades, crown age estimates still seem to be lacking.

We advocate a re-assessment of the effect of life-history and ecological traits on species richness, especially when species richness seems to be reduced. Sister clade comparisons, as well as other methods based on phylogenies, should take into account the influence of tree topology (species richness imbalances) and geometry (branch lengths) on the probability for transitions to occur: it is of great importance to know whether or not a trait is ancestral. Even while taking ancestry into account, sister clade comparisons can have a hard time identifying differences in diversification rates when transitions are common, leading to independent transitions in both sisters of a pair. In such cases, methods based explicitly on the phylogenetic tree, as BiSSE (Maddison et al. 2007; Fitzjohn et al. 2009) might be more useful. On the other hand, the statistical power of BiSSE decreases when one of the traits is rare (less than 10%, Davis et al. 2013); in such cases, sister clade comparisons might be the more useful tool. Recently, a method distinguishing transitions associated with speciation (cladogenetic) from those that occur along lineages (anagenetic) has been developed (Magnuson-Ford and Otto 2012), but its behavior in the
case of missing data and rareness of one of the character states is not clear yet. Even if future studies would find traits that are correlated with lower diversification rates, we warn against inference of a causal relation. Consider a trait without any effect: diversification could be reduced due to other factors that could be different in different parts of the tree of life. The longer branches that result as a consequence of these unmeasured traits could attract the measured trait changes to these clades.

SUPPLEMENTARY MATERIAL
Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.jd8vg.

FUNDING
This work was supported by the Agence Nationale de la Recherche (ANR-11-BSV7-013-03).

ACKNOWLEDGMENTS
We thank Gabriel Marais for discussions and suggestions for improvement of the manuscript; Thomas Bigot and Rich FitzJohn for help with simulations using BiSSE. We also thank Arne Moores, the editors Ronald DeBry and Luke Harmon, and two anonymous referees for valuable comments that helped to improve this manuscript.

APPENDIX
We compute the conditional expectation of a root edge length \( L \) in a speciation tree, given the bipartition resulting from the initial speciation event. The length of a root edge having \( p \) descendant leaves in a tree with \( m \) leaves is denoted by \( L_{m,p} \).

We use the Yule speciation model with a speciation rate \( \lambda \). In a Yule tree, speciation events occur at the internal vertices. The first speciation event is (rank 1) the root vertex and the waiting time between the speciation events of ranks \( i-1 \) and \( i \) is drawn from an exponential distribution with mean \( \frac{1}{\lambda} \). We choose an arbitrary orientation of the tree such that the shorter of the two edges descending from the root is on the left side.

The rank of the first speciation event in the right side of the tree is denoted by \( S(S \geq 3) \) because of the arbitrary orientation of the tree.

**Theorem 1.** The conditional distribution of the number of leaves in the right clade \( K \), given the total number of species in the tree \( m \geq 3 \) is

\[
P(K=k|M=m) = \frac{2(m-k-1)}{(m-1)(m-2)} \quad (1)
\]

**Proof.** The \( i \)-th speciation event (\( i \geq 3 \)) may occur on any of the \( i \) pending edges at that time with equal probability. The \( j \)-th (respectively \( r \)-th) speciation event on the left (resp. right) part of the tree may occur on any of the \( l \) (resp. \( r \)) pending edges in the left (resp. right) part of the tree with equal probability. The \( k \) (resp. \( m-k \)) leaves in the right (resp. left) part of the tree result of \( k-1 \) (resp. \( m-k-1 \)) speciation events. Because the second speciation event occurs on the left, there are \( \binom{m-3}{k-1} \) possible combinations for the ranks of the speciation events in the right (or left) part of the tree and the probability to obtain \( k \) leaves on the right of a tree with \( m \) branches is:

\[
P(K=k|M=m) = \frac{(m-k-1)(k-1)!}{(m-1)!} \times \frac{m-3}{k-1}. \quad (1)
\]

**Theorem 2.** The conditional probability that the longer root edge does not end before the \( s \)-th speciation event (\( 3 \leq s \leq m-k+1 \)), given the total number of leaves in the tree \( m \geq 3 \) and the number of leaves in the right part of the tree \( k \) is

\[
P(S \geq s|M=m,K=k) = \frac{m-s}{m-k} \frac{m-k-1}{(k-1)}. \quad (2)
\]

**Proof.** Using the same reasoning as for Theorem 1, if the first speciation on the right part of the tree does not occur before the \( s \)-th speciation event, then there are \( \binom{m-s}{k-1} \) possible combinations for the ranks of the speciation events in the right part of the tree and

\[
P(S \geq s|M=m,K=k) = \frac{P(S \geq s \cap K=k|M=m)}{P(K=k|M=m)} = \frac{(m-s)(k-1)!}{(m-k-1)!} \times \frac{m-s}{m-k} \frac{m-k-1}{(k-1)}. \quad (2)
\]

**Corollary 3.** In a Yule tree with a speciation rate \( \lambda \) and a total number of leaves \( m \geq 3 \), the conditional expectation of the length \( L_{m,k} \) of the longer root edge given the number of leaves \( k \) descending from this edge is

\[
E(L_{m,k}) = E(L'|M=m,K=k) = \frac{1}{\lambda} \left( \frac{1}{2} \sum_{i=1}^{m-k-1} \frac{(m-k-1)}{(k-1)} \right). \quad (3)
\]

**Theorem 4.** In a Yule tree with a speciation rate \( \lambda \) and a total number of leaves \( m \geq 3 \) the conditional expectation of the length \( L_{m,p} \) of a root edge given the number of leaves \( p \) in the tree descending from this edge is

\[
E(L_{m,p}) = E(L|M=m,K \in \{p,m-p\}) = \frac{1}{\lambda} \left( \frac{m-p-1}{m-2} \sum_{i=1}^{m-p} \frac{(m-p-1)}{(m-p-i)} + \frac{p-1}{2(m-2)} \right). \quad (4)
\]

**Proof.** The root edge leading to a tree with \( p \) leaves is either the longer (\( K=p \)) or the shorter root edge (\( K=m-p \)). The conditional probabilities of these two
cases given the total number of leaves are obtained from equation (Equation [1]). The expectation of the length of the shorter edge is \( \frac{1}{2} \) and the conditional expectation of the longer edge given \( K \) and \( M \) is obtained from equation (3).

REFERENCES


